

知识组织与知识管理

基于句子相似度计算的信息抽取*

廉站俊¹;吕学强¹;张玉杰²;施水才¹

北京信息科技大学中文信息处理研究中心¹

大连工业大学信息学院²

收稿日期 2007-5-10 修回日期 2007-5-10 网络版发布日期 2007-6-25 接受日期

摘要 提出一种基于句子相似度的信息抽取方法。采用句子主题相似度计算,对测试语料进行小句主题识别;同时结合句子主题在整个文章中的概率分布特点,提高识别的准确性。以网络上个人信息资源为语料,在该系统上进行测试,取得较好效果。

Abstract This paper gives a new method of information extraction based on calculation of sentence similarity. The topics of the sentences in testing words are labeled by adopting the method of calculation of sentence similarity. The veracity is increased by referencing the distributing of probability of the sentences in the documents. Using the resources of personal information on Internet, the paper achieves a statistic result.

关键词 [信息抽取](#) [概率分布](#) [主题](#) [句子相似度计算](#)

Key words Information extraction; Distributing of probability; Topic; Calculation of sentence similarity

分类号 [TP391](#)

DOI:

通讯作者:

廉站俊 dikk12345678@gmail.com

作者个人主页: 廉站俊 吕学强 张玉杰 施水才

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(431KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [参考文献\[PDF\]](#)
- ▶ [参考文献](#)

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [引用本文](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ [本刊中 包含“信息抽取”的 相关文章](#)
- ▶ 本文作者相关文章
 - [廉站俊](#)
 - [吕学强](#)
 - [张玉杰](#)
 - [施水才](#)